SNP Pipeline Commands

1. Index the reference genome using bwa index

```
/software/bwa-0.7.10/bwa index /reference/japonica/reference.fa
```

2. Align the paired reads to reference genome using bwa mem.
Note: Specify the number of threads or processes to use using the -t parameter. The possible number of threads depends on the machine where the command will run.

```
/software/bwa-0.7.10/bwa mem -M -t 8 /reference/japonica/reference.fa /reads/filename_1.fq.gz /reads/filename_2.fq.gz > /output/filename.sam
```

3. Sort SAM file and output as BAM file

```
java -Xmx8g -jar /software/picard-tools-1.119/SortSam.jar INPUT=/output/filename.sam OUTPUT=/output/filename.sorted.bam VALIDATION_STRINGENCY=LENIENT CREATE_INDEX=TRUE
```

4. Fix mate information

```
java -Xmx8g -jar /software/picard-tools-1.119/FixMateInformation.jar INPUT=/output/filename.sorted.bam OUTPUT=/output/filename.fxmt.bam SO=coordinate VALIDATION_STRINGENCY=LENIENT CREATE_INDEX=TRUE
```

5. Mark duplicate reads

```
java -Xmx8g -jar /software/picard-tools-1.119/MarkDuplicates.jar INPUT=/output/filename.fxmt.bam OUTPUT=/output/filename.mkdup.bam METRICS_FILE=/output/filename.metrics VALIDATION_STRINGENCY=LENIENT CREATE_INDEX=TRUE MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=1000
```

6. Add or replace read groups

```
java -Xmx8g -jar /software/picard-tools-1.119/AddOrReplaceReadGroups.jar INPUT=/output/filename.mkdup.bam OUTPUT=/output/filename.addrep.bam RGID=readname PL=Illumina SM=readname CN=BGI VALIDATION_STRINGENCY=LENIENT SO=coordinate CREATE_INDEX=TRUE
```

7. Create index and dictionary for reference genome

```
/software/samtools-1.0/samtools faidx /reference/japonica/reference.fa

java -Xmx8g -jar /software/picard-tools-1.119/CreateSequenceDictionary.jar REFERENCE=/reference/japonica/reference.fa OUTPUT=/reference/reference.dict
```

8. Realign Target

```
java -Xmx8g -jar /software/GenomeAnalysisTK-3.2-2/GenomeAnalysisTK.jar -T RealignerTargetCreator -I /output/filename.addrep.bam -R /reference/japonica/reference.fa -o /output/filename.intervals -fixMisencodedQuals -nt 8
```

9. Indel Realigner

```
java -Xmx8g -jar /software/GenomeAnalysisTK-3.2-2/GenomeAnalysisTK.jar -T IndelRealigner -fixMisencodedQuals -I /output/filename.addrep.bam -R /reference/japonica/reference.fa -targetIntervals /output/filename.intervals -o /output/filename.realn.bam
```

10. Merge individual BAM files if there are multiple read pairs per sample

```
/software/samtools-1.0/samtools merge /output/filename.merged.bam /output/*.realn.bam
```

11. Call SNPs using Unified Genotyper

```
java -Xmx8g -jar /software/GenomeAnalysisTK-3.2-2/GenomeAnalysisTK.jar -T
```

```
UnifiedGenotyper -R /reference/japonica/reference.fa -I /output/filename.merged.bam -o
filename.merged.vcf -glm BOTH -mbq 20 --genotyping_mode DISCOVERY -out_mode EMIT_ALL_SITES
```